

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 00:07:46 ; Search time 61 Seconds
(without alignments)
3381.199 Million cell updates/sec

Title: US-10-046-433-40
Perfect score: 5506
Sequence: 1 MAERGHSHLSARVGRTER.....LGRSNHLPKGLMDLQCR 1001

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2998	54.4	580	4 Q9P2M2	Q9P2M2 homo sapien
2	1350	28.2	493	4 Q96DP2	Q96DP2 homo sapien
3	1282	23.3	300	11 Q8R215	Q8R215 mus musculi
4	1203.5	21.9	1019	5 Q9NA40	Q9NA40 caenorhabdi
5	230.5	4.2	3567	11 Q9ES77	Q9ES77 mus musculi
6	211	3.8	709	5 Q97444	Q97444 giardia lam
7	208.5	3.8	719	5 Q9U019	Q9U019 giardia lam
8	208.5	3.8	1101	5 Q964D2	Q964D2 entamoeba h
9	207	3.8	709	5 Q9XTU7	Q9XTU7 giardia lam
10	206	3.7	1007	13 Q902N3	Q902N3 gallus gall
11	202.5	3.7	719	5 Q9U021	Q9U021 giardia lam
12	200.5	3.6	1316	4 Q96JU7	Q96JU7 homo sapien
13	197.5	3.6	1045	5 Q8T3A6	Q8T3A6 caenorhabdi
14	197.5	3.6	1070	5 Q8T3A7	Q8T3A7 caenorhabdi
15	197.5	3.6	1111	5 Q9XWD6	Q9XWD6 caenorhabdi
16	195.5	3.6	1074	5 Q964D1	Q964D1 entamoeba h

17	195	3.5	2112	5 Q8WPL0	Q8WPL0 olkopleura
18	188.5	3.4	1623	5 Q93037	Q93037 anopheles g
19	187	3.4	1875	5 Q93691	Q93691 caenorhabdi
20	186	3.4	704	5 Q9U048	Q9U048 giardia lam
21	185.5	3.4	556	5 Q9NC23	Q9NC23 giardia lam
22	183.5	3.3	1391	5 Q19021	Q19021 caenorhabdi
23	182.5	3.3	769	5 Q24971	Q24971 giardia lam
24	181.5	3.3	667	5 Q95WU1	Q95WU1 giardia lam
25	181	3.3	667	5 Q9XWU1	Q9XWU1 giardia lam
26	179	3.3	1274	5 Q9NCL3	Q9NCL3 giardia lam
27	178	3.2	1274	5 Q24977	Q24977 giardia lam
28	177.5	3.2	1193	13 Q90819	Q90819 gallus gall
29	177.5	3.2	2534	5 Q8WMD1	Q8WMD1 plasmodium
30	177	3.2	557	5 Q24992	Q24992 giardia lam
31	175	3.2	2352	5 Q61240	Q61240 halocynthia
32	174	3.2	3386	5 Q9VM55	Q9VM55 drosophila
33	173.5	3.2	1290	5 Q9VTR8	Q9VTR8 drosophila
34	172	3.1	1154	5 Q9G046	Q9G046 giardia lam
35	172	3.1	1372	5 P91526	P91526 caenorhabdi
36	171.5	3.1	739	5 Q9GS24	Q9GS24 giardia lam
37	171	3.1	2447	5 Q9NMF9	Q9NMF9 drosophila
38	171	3.1	4072	5 Q9M4Y4	Q9M4Y4 drosophila
39	170.5	3.1	1214	13 Q90YD2	Q90YD2 xenopus lae
40	170.5	3.1	1324	5 Q85YK2	Q85YK2 drosophila
41	170	3.1	2759	5 Q45614	Q45614 caenorhabdi
42	170	3.1	3102	5 Q9TZR4	Q9TZR4 caenorhabdi
43	169	3.1	1376	5 Q85ZS2	Q85ZS2 drosophila
44	169	3.1	1679	5 Q24301	Q24301 drosophila
45	167.5	3.0	545	5 Q9G044	Q9G044 giardia lam

ALIGNMENTS

RESULT 1	Q9P2M2	PRELIMINARY;	PRT;	580 AA.
ID	Q9P2M2	Q9P2M2		
AC	Q9P2M2	Q9P2M2		
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DE	K1A1324 protein (Fragment).			
DE	K1A1324 protein (Fragment).			
GN	K1A1324.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RX	MEDLINE=20181126; PubMed=10718198;			
RA	Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;			
RT	"Prediction of the coding sequences of 150 new cDNA clones from brain which code			
RT	for large proteins in vitro.";			
RL	DNA Res. 7:65-73(2000).			
DR	EMBL: AB037745; BAA92562.1; -			
FT	NON_TER			
SQ	SEQUENCE 580 AA; 63365 MW; 5E63C19265EC8E5C CRC64;			
Query Match	54.4%; Score 2998; DB 4; Length 580;			
Best Local Similarity	99.1%; Pred. No. 7.6e-236;			
Matches	555; Conservative 1; Mismatches 4; Indels 0; Gaps 0;			
QY	425 WNTLPLNNMETVYVSGINFEYKGMGTGEVAGDHITTYRAGASDNDPMILTYVGFPPPOS 484			
DB	1 WNTLPLNNMETVYVSGINFEYKGMGTGEVAGDHITTYRAGASDNDPMILTYVGFPPPOS 60			
QY	485 VMADTEKEKVARITFVEETLCVNCLEYEMVGVNSRTPVEVFWKSKGKOSTYIIIEEN 544			
DB	61 VMADTEKEKVARITFVEETLCVNCLEYEMVGVNSRTPVEVFWKSKGKOSTYIIIEEN 120			
QY	545 TTTSFTWAFORTFHEASRKYTNDVAKIYSINVTNMGVASYCRCALEASDVGSSTCTS 604			

```

|||||
Db 121 TTSTSTWAFQRTTFHEASRKYNDVAKISINTVNMNGVASVCRCALEASDVSSCTS 180
OY 605 CPAGYIIDRSDSTCHSCPPNTILKAHQPYGVACVPCGGTGNKNTIHSCTYDCFTSRNT 664
Db 181 CPAGYIIDRSDSTCHSCPPNTILKAHQPYGVACVPCGGTGNKNTIHSCTYDCFTSRNT 240
OY 665 PRFTFNYSALANTVTLAGSPFTSGKGLKYPFHHTLSLCGNGRKMSTCTNNVTLRLP 724
Db 241 PRFTFNYSALANTVTLAGSPFTSGKGLKYPFHHTLSLCGNGRKMSTCTNNVTLRLP 300
OY 725 EGSEGSFKSTAYVCOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTMDLTDLGITSAPL 784
Db 301 EGSEGSFKSTAYVCOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTMDLTDLGITSAPL 360
OY 785 FHLESLGIPDVIFFYRSNDVYQSSSGRSTTIRVRCSPKTYPGSILLPFGTSDGICDCG 844
Db 361 FHLESLGIPDVIFFYRSNDVYQSSSGRSTTIRVRCSPKTYPGSILLPFGTSDGICDCG 420
OY 845 NFHFIMESAACPLCSVADYHAIVSSCVAGIQKTYVWREPKLSCGSLPDRQVYICKT 904
Db 421 NFHFIMESAACPLCSVADYHAIVSSCVAGIQKTYVWREPKLSCGSLPDRQVYICKT 480
OY 905 IDFWLKVGISACTCAILLTVLTCYFWKKNOLEKYSKLVNNAATLKDCDLPADSCALM 964
Db 481 IDFWLKVGISACTCAILLTVLTCYFWKKNOLEKYSKLVNNAATLKDCDLPADSCALM 540
OY 965 EGEDVEDDLIFTSKNHSLGR 984
Db 541 EGEDVEDDLIFTSKNHSLGR 560

```

RESULT 2

```

OY 096DP2 PRELIMINARY: PRT: 493 AA.
AC 096DP2:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CDNA FLJ131340 f1s, clone MESANI000035, weakly similar to major
DE surface-labeled trophozoite antigen precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagatsuma M.,
RA Kawakami K., Kanehori K., Takahashi-Fuji A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K.,
RA Isogai T.;
RT "NDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK055902; BAB71041.1; -.
DR InterPro; IPR001881; EGF_CA.
DR SMART; SM00179; EGF_CA.1;
SQ SEQUENCE 493 AA; 54585 MW; 0AEBB2C19709B211 CRC64;

```

Query Match 28.2%; Score 1550; DB 4; Length 493;
 Best Local Similarity 55.7%; Pred. No. 7.3e-118;
 Matches 272; Conservative 80; Mismatches 132; Indels 4; Gaps 4;

```

OY 231 HSEVLEGNVNVYRTTFVSVTKPPLVLAITGAAYVSECFPCPGTYADKQSS 290
Db 6 YQVMLKSGNIIYKRTTGILMGSKAVKPLVLAITIEGAYVSECFPCPGTYADKQSS 65
OY 291 FCKLCPANYSYKSGTSHOC-DPDKYSKSGSSCNVPRPACTDKDYFTYHTACDANGETO 349
Db 66 NCOVCPRNTYSEKGAKECIRCKDSDQFSEGSSECTERPPCTTKDYFOIHPCDEBGKTO 125

```

```

OY 350 LMYKNAKPKICSEDLGCAVKAFLPASGVKTHCPNCPGFETKNNSTQCPYGSYNGS-DC 408
Db 126 IMKWIEPKICREDLTDALRLPSSGEKKDCPCPCNPBFYNNSSSCHPCPPGTFSDDTKRC 185
OY 409 TRCPACTEPAVGEYKWMNLTPTNMTTVLSGINFYEKMTGWEVAGDIIYTAAGASND 468
Db 186 RCPACTEPAVGEYKWMNLTPTNMTTVLSGINFYEKMTGWEVAGDIIYTAAGASND 245
OY 469 EMILTIYVGEFRPQSVAMDTEKKEVARTTFVETLCSVNCLEFVWVGNSTRNPFVEN 528
Db 246 YILNLNHIIGFKRPTV-MTGATGSELGRITTFVETLCSADCVLEFVWDINRSTNVESM 304
OY 529 KSKSGKOSTYLIIEENTTSFTWAFQRTTFHEASRKYNDVAKISINTVNMNGVASVCR 588
Db 305 GGTKEQATTHIIFKNAITFTWAFQRTNOGDNRFRINDMKIYISITATNNAVGVASSC 364
OY 589 RCPALASDVSSCTSPAGYIIDRSDSTCHSCPPNTILKAHQPYGVACVPCGGTGNKNT 648
Db 365 RCPALASDVSSCTSPAGYIIDRSDSTCHSCPPNTILKAHQPYGVACVPCGGTGNKNT 424
OY 649 KIHSLCYNCTSRNPTFTFNYSALANTVTLAGSPFTSGKGLKYPFHHTLSLCGNG 708
Db 425 ODHAVCYSDCFEYPERKENOSLHDFSNLSSVGLMNGPFTSGKTYEFHFNFISTYVGM 484
OY 709 RK-MSVCT 715
Db 485 RKMLSVPT 492

```

RESULT 3

```

OY 08R215 PRELIMINARY: PRT: 300 AA.
AC 08R215:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 32.6 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strusberg R.;
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC022655; AAH22655.1; -.
FT Hypothetical protein.
FT NON_TER
SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;

```

Query Match 23.3%; Score 1282; DB 11; Length 300;
 Best Local Similarity 86.0%; Pred. No. 2.6e-96;
 Matches 233; Conservative 16; Mismatches 22; Indels 0; Gaps 0;

```

OY 714 CTDNVTDLRIPEGESEFSKSTAYVCOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTM 773
Db 1 CTDNVTDLRIPEGESEFSKSTAYVCOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTM 60
OY 774 TLDGTSPELPHLESLGIPDVIFFYRSNDVYQSSSGRSTTIRVRCSPKTYPGSILLP 833
Db 61 TLDGTSPELPHLESLGIPDVIFFYRSNDVYQSSSGRSTTIRVRCSPKTYPGSILLP 120
OY 834 GTCSDGTDCGCFHFLMESAAACPLCSVADYHAIVSSCVAGIQKTYVWREPKLSCGITS 893
Db 121 SMCSDGTDCGCFHFLMESAAACPLCSVADYHAIVSSCVAGIQKTYVWREPKLSCGITS 180
OY 894 LPEQRTYICKTIDFWLKVGISAGTCAILTVLTCYFWKKNOLEKYSKLVNNAATLKDC 953
Db 181 LPEQRTYICKTIDFWLKVGISAGTCAILTVLTCYFWKKNOLEKYSKLVNNAATLKDC 240
OY 954 DLPADSCALMEGEDVEDDLIFTSKNHSLGR 984

```

Qy	49	CHESEHYEYTCADSTGSRMYAAVPHRPGLOSTLDPDKVKECTSEFCNGAEJDMKDDSC	108
Dd	25	CTDDDELYEYTCMDENGERMYAVPRGMCQSNLPTFRRLGNCFSFCDEPHYLDLDSHC	84
Qy	109	KPCAEGRYSLGTGIRFDEMDLPHGFA--SLSANMELDDSAAES-IGNC-TSSKVPBRD	164
Dd	85	RPCNPGFFSLGGIRHEEVEVLTPSGFSVDNNDSDNDAOFNSRDSQVECKEKGWVYKCG	144
Qy	165	YIAFNTDECTALTMVAVNLKSGGTAFNYEYYP--DSIIIEFFEQNDQCPNAD--DSRW	220
Dd	145	ELIYIDPTPCVSKLSEFANLYVRGSAVEFYRMPRRNRRLASMQVDLRNQCOSYDVAKSMF	204
Qy	221	MKTTER-----GMEHSVELRGNVNLVWRTTASVWTKYKPYVLRNLTATGVAY	271
Dd	205	LKTKKKEDEEKNQDMRRRIELKSGANVLSWIIIONNMGOASNOPIHIDRIDVJGLAF	264
Qy	272	TSECFPCPKCTYADKOGSSFCFLCPCANSYKNGETSCHCQDPDSEKSGSSCNCVRPACH	331
Dd	265	TROCTACPEPT-SSPGGSAECTIPCSBGSFSSKSGSQCCRCDESGYSGFKSEKCIDRPPCR	323
Qy	332	DKDFTYHTACDANGETOLMYKMAKPICTCEDLSEGAVKLPAISGVKTHCPCPNCGEFTTN	391
Dd	324	VSQDYPAVREPC-INGSSRAVYKVKVLPISICRDPLPATKILPPTPMKTCPCNPGEMENKL	382
Qy	392	STQPCPYRGYSNGSOTCOTPAQTEPAVGEFEKMMNNTLPTMMET-----TVLSGJNFE	444
Dd	383	GVCFECCKKDYFSDGNSCRQPDVTPYNTGLQYQNMENVLPKLSTRCEYISBDVATACNIG	442
Qy	445	YKGMTGMEVAGDHIYTPAAGAS-DNDEMLTLVVPGF-RPPOSVMADITENKVALITYFE	502
Dd	443	-----DAMIWPGSDSLISAPSLTGIAIELLISIDEGFWMLPARKPKTKYVAVAAVITYFE	498
Qy	503	TLCVS-VNCELTYEMGVGSKRNTPEV-----WKSCKKOSYTYIIEENTTSFTSAFO	554
Dd	499	TSCADESALCFEIDMSAGIKGORESYHEFLAIFNCSNKRKWSMTITVTKNTPAPFMAFL	558

AC Q95877 01-MAR-2001 (TrEMBLrel. 16, Created)
AD 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Polydorm protein precursor.
GN POLYDORM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HENSNC;
RX MEDLINE=20517255; PubMed=11062057;
RA Gilles D., Vanit M.-A., Callebaut I., Coulombel L., Cacheux V.,
RT "Polydorm : a secreted protein with pentaxin, complement control
RT protein, epidermal growth factor and von willebrand factor A
RT domains".
RL Biochem. J. 352:49-59(2000).
RL EMBL; AF206329; AAC32160.1; -.
DR HSSP; P00740; IEDM.
DR MGD; MGI:1928849; Polydorm.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003410; HyalIn.
DR InterPro; IPR001759; Pentaxin.
DR InterPro; IPR000436; Sush_SCR_CCP.
DR InterPro; IPR002035; WVF_A.
DR Pfam; PF000008; EGF; 10.
DR Pfam; PF02494; HXR; 2.
DR Pfam; PF00354; pentaxin; 1.
DR Pfam; PF00084; sush; 33.
DR Pfam; PF00092; wva; 1.

DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00895; PENTAXIN.
 DR PRINTS; PR00453; VMEFAD0AIN.
 DR PRODOM; PD002153; Pentaxin; 1.
 DR SMART; SM00032; CCP; 34.
 DR SMART; SM00181; EGF; 15.
 DR SMART; SM00179; EGF_CA; 9.
 DR SMART; SM00001; EGF_1like; 3.
 DR SMART; SM00159; PTX; 1.
 DR SMART; SM00327; VMA; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 5.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_9.
 DR PROSITE; PS01186; EGF_2; 11.
 DR PROSITE; PS01187; EGF_CA; 6.
 DR PROSITE; PS50234; VMEF; 1.
 FT SIGNAL.
 KW SIGNAL.
 SQ SEQUENCE 1 17 POTENTIAL.
 3567 AA; 387391 MW; 87BA8276E12293B5 CRC64;

Query Match 4.2%; Score 230.5; DB 11; Length 3567;
 Best Local Similarity 20.4%; Pred. No. 4.5e-09;
 Matches 174; Conservative 99; Mismatches 300; Indels 279; Gaps 46;

QY 47 HACKESYHEHYEACSTGSRMAYAPHPGCLCTSLDPVKGTECSFGSCNAGEFLD--MK 104
 DB 795 HGFSEDMKKTTRCDMDLFFKESAFETTLGNMVP-----SF-CNDADDIDICRL 845
 QY 105 DOSCKPCAGRYSL-----GTGIRFD-EMD-----ELPHGF-----ASLSAN 140
 DB 846 DLTKKCYIEYNNYENGFAIGGCMAGNRDIYSDHFLDVOEFTPDYAKRSSRIKPT 905
 QY 141 MELDSDAESTGNCSTSSKVP--RGDYAFNDECTATLMVAVNLKSGTVNFEYYPPDS 198
 DB 906 VPLSDPKIOLINIVASVPLPEERNDTLELNOQRILIKTLFTITRLKSLINKEMYS-- 963
 QY 199 STIFFEYVONDCOPNADSRMMKTEKGMEFHSVELNGNVLWRTTASVTKVPR 258
 DB 964 ---FQASFTVADNSLE-----TEKAF-----LECR-----PGS 991
 QY 259 VLVNRNAT--GVAYT--SECFPKQGYADKQSSFCILPANYS--NKGETSC 308
 DB 992 VLRGRVCVCPLGTSYSLHSTCESCLMGSYQDEEGQLECKLCPRRTHEVLHSSVSEC 1051
 QY 309 H-QCDPDKYSEKSSGSSCWVRPACTDKQFYHTNCADANGCFQIMAKKPKICSEDEGA 367
 DB 1052 KAQCKQGGYSSSGLETSCSLPTGYQPFGRS-SCLLCEPTTTTVK-----RGA 1099
 QY 368 VKLPASGVKTHCPNCPNGEFTNNST-CQPCP--YGSYNSGSDCTRCRPAGTEPAGVPEY 423
 DB 1100 VDISAGV---PCPYGEFSRSGILTPCYPCPRDYQPNACKSFCILACP----- 1143
 QY 424 KMMWTLPTNMTETVLSGNEFYKMTGMEVAGDHI--YTAAGASDNDPMI----- 471
 DB 1144 -FYGTITTTGATSTITDCSSFS--STFSAEESIVLVAAPHGSHQNKYEVSQVHFECFL 1198
 QY 472 -----LTLVVGFRPPQSVADTENEKEVARITPVEFTLC-----SVN 508
 DB 1199 NPGHNSGTCOOLGRGYVCLCPGV---TGAKCEFDIDECSSILPCNLNGICRDVOYGCPTCE 1255
 QY 509 CELFPMGVNSRTNTPETWKGSGKQSYIIEENTTTSFTWAFORTTTPHEASRKTYND 568
 DB 1256 CSL-----GYSGQ-----ICEEN-----INE 1271
 QY 569 VAKIYSINVTNVMGVASVCRPCALASDVGSSTGSCPNXY--YIDROSGTCHS--CPP 623
 DB 1272 CISSPCLNKTCTDGLASY-----RGT-CVKGVMGVHCEETDNEQSSPCSLN 1317
 QY 624 NTILKAHQPIGVOA-CVPCGPGTKNNK-----IHSLYNDCTFSRMTPTRENYNSALA 677
 DB 1318 NAVCK-DQVGFSCKCPRGFLGTRCKENVDDECLSQPCQNGATCKDG-----A 1363
 QY 678 NTYTLAGGPFSTKGLKYHFHFTLSL--C-GNGGRKMSVCTDNVTDLRIPEGSEGSKSI 734

DB 1364 NSERCOPAGFTGT-----HCELININEQSNPCRNQATCVD-----EL 1401
 QY 735 TAYCOAVIIPPEVTGYKQVSSOP--VSLADRLGVTITDMLDGTTSAPAEFLHESLGI 792
 DB 1402 NYSCKC---QPEFSGHRCE-ETEDPSGFNLDPREVSGITGYVLLDGV-----LPTLHA 1449
 QY 793 PDVIFPFRSNDV 804
 DB 1450 ITCAFMKSSDV 1461

RESULT 6
 097444
 ID 097444 PRELIMINARY; PRT; 709 AA.
 AC 097444;
 DT 01-MAY-1999 (TREMBLrel, 10, Created)
 DT 01-MAY-1999 (TREMBLrel, 10, last sequence update)
 DT 01-JUN-2002 (TREMBLrel, 21, last annotation update)
 DE Variant-specific surface protein 417-4 (Variant-specific surface protein type 4 TSP11/TSA417-like).
 GN VSP417-4.
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
 OX NCBI_TaxID=5741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AD-1;
 RA Ey P.L., Darby J.M., Mayrhofer G.;
 RT "A new locus (VSP417-4/A-I) belonging to a subfamily of tsaa17-like variant-specific surface protein (VSP) genes in Giardia intestinalis.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 476-705 FROM N.A.
 RC STRAIN=VARIOUS STRAINS;
 RA Ey P.L., Darby J.M.;
 RT "Identification of a subset of tsaa17-like genes within the variant-specific surface protein (VSP) gene family of Giardia intestinalis.";
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF065168; AAD0439.1; -;
 DR EMBL; U08153; AAD0504.1; -;
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR005127; Giardia_VSP.
 DR Pfam; PF03302; VSP; 2.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00001; EGF_1like; 3.
 DR SMART; SM00261; FU; 5.
 SQ SEQUENCE 709 AA; 72664 MW; 6E71F27D2F367F52 CRC64;

Query Match 3.8%; Score 211; DB 5; Length 709;
 Best Local Similarity 20.7%; Pred. No. 1.9e-08;
 Matches 151; Conservative 65; Mismatches 273; Indels 242; Gaps 36;

QY 4 PGHSHLSARVNRTERIRIPRLRLLMAGTAFOYTGTPGLHACKSEHYEYACDS 63
 DB 99 PGHSLCISDGDGVCTEAPGYFLNPLRANTKDSV-----SCSDT-----TGFTD 144
 QY 64 TGSRRMVAVPHPTPGCLTSLDPVKGTECSFGSCNAGEFLDMK--DOSCKPCAGRY----- 116
 DB 145 SKGTYR-----GVQYCERCDAALTDAGGDGAKKRCRCODKYLKNDT 186
 QY 117 -----SLGTGIRPDEMDELPHGFASISANMELDASAESNGTSSKWPPRDGYAFN 169
 DB 187 CYDKAQCSGSGTNKRVAYADDSENGKNKYSCSNLNGVA---NCDTCSYDQSKRTI-- 239
 QY 170 TDECT-ATLMVAVNLKSGTVNFEYYPPDSITIEFFVONDQCPN--ADDSRMKATTE 225
 DB 240 -KCTKCTDNNVLTKTTSBGT-----SCVQKQCKDGDGFPKXDS----- 275
 QY 226 KGMFHSVELNNGNVLWRTTASVTKVPRVIVRIAL-----TGVAYISECPFC 278
 DB 276 -----SAGKCLCPNDSTDG-----IANCATCALVSGRSGAALVT-CSAC 314

KW Lectin. 1101 AA: 119512 MW: CGB6F5CDBDE656AEC CRC64;
 SQ SEQUENCE

Query Match
 Best Local Similarity 18.3%; Score 208.5; DB 5; Length 1101;
 Matches 229; Conservative 113; Mismatches 396; Indels 501; Gaps 61;

QY 57 EYACDSTGSRM-RVAVPH-----TPGLCTSLP----- 83
 DB 14 DYADKLIGKEPREAVPHCASVSNAGCTSDTGYLTGTTGNNKCTTLEKDKCTAFSYX 73
 QY 84 DPVWGT--ECSFSCNAGEF-----LDMKDGSCPKCAGGRSL 118
 DB 74 DKRNSTNPKCTCYVNGKEVNTSSHSNDKCYCKNNVNICESLMLKKDKSCGCIIGMSTT 133
 QY 119 GTGIREDEMDLPHGFASLSANMELDDSAESTG--NCTSKWVPKRDYIAFNTDECTA 175
 DB 134 VDG-----SKLCDNATEDNAENKCVGLLASTTSK-----TCDKCFG 170
 QY 176 TLMYAVN-----LKSGTVFEYIYPPDSIIFFEVQNDQCPNADDSRPMKTEKGMEFH 231
 DB 171 --MYSLGSGKCTOKNDKIN-----KCILOVENSQCADGYSLSIDKSKAKPE-H 219
 QY 232 SVELNGNNVLYWRTTAFSVMTKVPKPYLVNIAITGVAITSECPCKPGTYADKOGSSF 291
 DB 220 CSKIN-GN-----OCLTMEGYLSKETDSK- 243
 QY 292 CKLCPANSYKNGETSCQCPDKYISEKSSCNVRACTDKDYFYHTACDANGETOJM 351
 DB 244 CTICVDPNPNLSRG--NEC--SIYNAHCTSKNR--CTVSGVCYCKNNCRIFSEFTE-- 295
 QY 352 YKMAKPKICSEDLGAVKLPASGVKTPCPCPNNGEFTKNTSTQCPKPGSY--SNGSDC- 408
 DB 296 -----ENKCTKCDMGFLITTSIGTSPNLYOGFKTANRTECE 331
 QY 409 -----TRCPACTEPVAVFEYKWMNTLPTNMTETVLGIN----- 442
 DB 332 NGVYLEDDGKKRCSLCPDFTCLTKSKTPVPQ-----KLNLRSHLSTGPKCL 382
 QY 443 -----FEYKGTGMEVAGDIY-----TAGASDNPMLITLVPGFRPQSVAM 486
 DB 383 PGLCLSDDDTTCYKENGTLNGTCHYFNDRKSVLGISGNHGVCKM--RQYDYEYOYL 440
 QY 487 ADENKEVARITFEFTLCSVNCLELYPMGVNSRTNPTVEWKGSKGOSYTIIEENTT 546
 DB 441 -----NAFKASDNTYTCPLDLPLPYFYSY-----TKG-----TSDNTI 474
 QY 547 TSETMAFQRTFHEASRKTYTNDVAKIYSINTVNNVNGVASYCRCALASAVSSCTSCP 606
 DB 475 TIGCVQGLNVSNDCE--CNDKHIPTSIDKASDCVSITITKLPSCERTAN--GNICTQCP 529
 QY 607 AG-----YIYDRDSCGCHSCPNTILKAHQPYGVQ--ACVPC-----GPGTK 646
 DB 530 VQSHGKGKCKSCGDHAFYFDKN--VCKKCPASCSSCSYDSKSKVYVSECYENIQTTR 588
 QY 647 NKKHSLCYND-----CTFSRMTPTRTFNYNFS-----A 675
 DB 589 NKENECACINDYKREKPAEDKKKSCAQLNNCKKBEKTEISGFEVYCLDDCDSAYIVGS 648
 QY 676 LANVTYLAGPSF-----TSKGLKYFH-----HFTSLICNGQRKMS 712
 DB 649 QVGACTQCSPAFENKNNKQOLCSTKQSYGHCACSAFACINCEIDINILITGE--KPCGT 706
 QY 713 VCTD-----NVYDLRIPEGESGFSKITAYVQAVIIPPEVYIAGVYSSQPSVLAIDL 766
 DB 707 VCKGQFOIENATD-----GYVCSPC-----PAKCKCTKYNTTSKRVCEV-- 746
 QY 767 IGVTTDNTLGITS-----PAELFHLIESLGIDPVIFFYRSNDVOTSCS----- 809
 DB 747 --TCTEORLDKTIKAPCACTGTVOLE-----NCTQCSGDLSTKYPCKCKTD 791
 QY 810 -----SGRS-----TTLIRVCSPOKTYVPSGLLIPG-----T 835

DB 792 SCNWDSTRTGLIYATECSDFSGRSRPSYNCCTT-----CTKSNYTPKEGKNCACKODKCAT 847
 QY 836 CSD-----GTCDGCGNFHFLM-----ESAACPLCSVADYHA 866
 DB 848 CSDKDTOLTCADPLKVGSKCDCKGTGYMSNGECKPCTNHCESSSAACCTVCESDTYKV 907
 QY 867 I-----VSSCYAG-----IQKTYVMREP-KLCSG-----GISLPEQRYV--TIC 902
 DB 908 ISGNGCNSCVDFGFPEIKGTCTIPCTSPCTKCVYKACDCEDETCNSBKRTVEECTKG 967
 QY 903 KTIIDFWLKVIGISAGCTCTAILLYLVCYEW-----KKNOKLEYKSKLYMN-----ATLK 951
 DB 968 STFDHIAEYVNV-GAC-----VCAITGVETSTEDNKIECQACKAKVNEFDCSCK 1018
 QY 952 DCLPLPAADSCALMEGDEVDLFTSK-NHSLGRNHLR 989
 DB 1019 DCLRCNAEYLEAKGCEYVCEGYTSSWGSCIPCRRHMP 1057

RESULT 9

09XTJ7

AC 09XTJ7 PRELIMINARY; PRT; 709 AA.

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Variant-specific surface protein (fragment).
 GN VSP417-6.
 OS Giardia lamblia (Giardia intestinalis).
 CC Eukaryota; Diplomonadida; Hexamitidae; giardinales; Giardia.
 NX NCBI_Taxid-5741;
 RN [1];

RP SEQUENCE FROM N.A.
 RC STRAIN-BRIS-136;
 RX MEDLINE-99026095; PubMed-9806870;
 RA Ey P.L., Darby J.M.;
 RT "Giardia intestinalis: conservation of the variant-specific surface protein VSP417-1 (TSA417) and identification of a divergent homologue encoded at a duplicated locus in genetic group II isolates."
 RL Exp. Parasitol. 90:250-261(1998).
 DR EMBL: U89266; AAD03483.1;
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR005127; Giardia VSP.
 DR Pfam: PRO3302; VSP_2; Kazal_inhib.
 DR PRINTS: PR00290; KAZALINHTR.
 DR SMART: SM00261; FU; 3.
 FT NON_TER
 SQ SEQUENCE 709 AA: 71516 MW: 3512BB844B38D134 CRC64;

Query Match
 Best Local Similarity 3.8%; Score 207; DB 5; Length 709;
 Matches 149; Conservative 70; Mismatches 280; Indels 244; Gaps 35;

QY 60 ACDSTGSRMRAVPHPTDELCTSLDPYKGTGCSFSNA-----GEPLDM----- 103
 DB 18 ACQADGS-----GSAGSCKTCGVALIGQPYCECGANGAYAVVNOQADVNAECPSTKL 70
 QY 104 ---KDO--SCKPC-----AEGRSIGTGIRFEMDELPHGFASLSANMELDDSAESTG 152
 DB 71 CPARDQCKKTCQGGNSMLAGCCYSSGEG-----LPGHSLICLSSG-----GDG 113
 QY 153 NCTSKWVPKRDYIAFNTDECTATLMYAVNLKOSTGVNEEYVY-----POSIIFEEFYOND 209
 DB 114 VCTEAPAGYFLNPLRANTKDSVYSCSDTLAGFTDSKTYGVOYCERCBAVLTDAAGDA 173
 QY 210 QCPNADDSRPMKKT-----EKWGFHVELANGNNVLYWRTTAFSVMTKVPKPYLVR 262
 DB 174 KC--TRGEMNKYIATITGCEGCTPTFEFSKEDSDNG-----KRCFAC 214
 QY 263 NAITGVAITSECFCKPCTYADKOGSSFPCKLCPANSY-----SNGGTSCHO--CDPDKY 316
 DB 215 GDVTTGAVASCERCTPPSP-----DAKPACTKCGGNNILKTAADGTYTCAEQSACSPDSF 269

```

Oy 317 SEKSSSSCN-----VRPACTDKDY-----FYTHAC-----342
Db 270 EVENSQSGNRCVLCGDAANGVDKCACTPADKRAAPAVTCTACTDGYKPSADKTCEA 329
Oy 343 -----DANGER-----OLMTKAKAPKICSDLEGAUKL---370
Db 330 VSSCTPCCKACSNKGENEVCTDCDSYLTPTSOCIDSCAAGVNYGALGAKLCKE 389
Oy 371 -PASVKT-----HCPPNPFEEKTNNSTOCPCPYG-----SYNGSDCTRCAPGEPAYG 420
Db 390 CTAAANKCTCDGGRCQCTSDSFYK--NGDACSPECHESKCTKSAGTADCTCEPFG-----442
Oy 421 FEKMMNTLPTNMTTYSVLSGINFEXKMTGWEVAGDHITAAASDNDEPILTLVYRGR 480
Db 443 -----KALRYGDDGTGKTCGAG-----CATGSGSACACTGCLTLIDGAS 480
Oy 481 -----PPQVYADTENK--EVARITFEETLCSVCELXYFW--GVNSRTNTPV 525
Db 481 YSECATATEYEPONGVCAKASRATPTCNDSPIONGVCGICANCFKMGCGEYVYKPG 540
Oy 526 ETW-----KSGKQSYTYIIIEENTTSFTWAFQRTTEHASKRYNDVAKIYINVTN 579
Db 541 KTYCISAPNGCTQKADYKIDSGTL-----TVCSEGCCECTS-----STDCIT 585
Oy 580 VMNGV--ASYCRPCALE--ASDVSSCTSCAPAGY-IDRDSGCHSCPPNTILKAHP 632
Db 586 CLDGYKASACYKCDPSCETCGAATTCACATGYITKTAGSEGACTSCESD-----SNGV 641
Oy 633 YGVQACVPCGPGTKNNKIHSLCY 655
Db 642 TGKGLNCAPPPSSNTG-SVLCY 663

RESULT 10
O902N3 PRELIMINARY; PRT; 1007 AA.
AC 0902N3.
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Laminin gamma 1 (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RA Halfter W., Dong S., Balasubramani M., Bier M.E.;
RT "Aberrant histogenesis after temporary disruption of the retinal basal
lamina."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF373841; AAK55397.1; -.
DR InterPro: IPR004089; Chmlaxis_transd.
DR InterPro: IPR000361; EGF-like.
DR InterPro: IPR000344; Laminin_B.
DR InterPro: IPR02049; Laminin_EGF.
DR InterPro: IPR001866; LamNT.
DR Pfam: PF00052; laminin_B_1.
DR Pfam: PF00053; laminin_EGF_10.
DR Pfam: PF00055; laminin_Nterm; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_6.
DR PROSITE: PS01186; EGF_2; UNKNOWN_6.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; UNKNOWN_9.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1007 AA; 110999 MW; CCBFD9659EF93JPC CRC64;

Query Match 3.7%; Score 206; DB 13; Length 1007;
Best Local Similarity 20.6%; Pred. No. 7.8e-08;
Matches 165; Conservative 80; Mismatches 309; Indels 248; Gaps 46;

```

```

Oy 209 DQCPNADDSRMK--TTEKGEHSHVELNRGNVLYWRTAFSVWTKVPVLVNIAT 267
Db 72 EKLPFFNDPRWRATATASANECLPCDNGRSDECT-----DPELYRS---T 116
Oy 268 GVAIYSECPCK--PGTYADKQSSP-----CKLCPANYSKN-----GETS--307
Db 117 G--HGCHMGCGDNTDGAHCERCNRDSFYRLGSEGCJPCSCNPVGSISTQDSTYGCCSK 174
Oy 308 -----CHQCDP--DKYSEKSSSCNVRPACTDKDYFTHACDANGETOLMYKAKRK 358
Db 175 PGVMEKCDRCOPGFHSLSEACRCPCSCNPA-----GSTGECHMETGRC-----218
Oy 359 ICSEDEGAVKLPASGVKHCPCNPGFE--KTNNSTOCPCPYGSTSKNSDCTRCAGT 415
Db 219 TCRDNVGE-----FHCRCRKGFEHLDPSPNRCPTPC--FCGHSSVCTN-----261
Oy 416 EPVAFEYKMMNTLPTNMTTYSVLSGINFEXKMTGWEVA--GDHITAAASDNDEPIL 472
Db 262 --AVGYSV-----YSITSPFO--GEDEHABQRDGSGVPLQWSEETQDISVI 305
Oy 473 T-----LVVPGRRPQVYADTENKEVARITFEET--LCSVCELTYMVGVNSRTN 522
Db 306 SDSYFPIYVAPRKFLGNVLSYGQ-----LTFSPVRDRDTRLASADVLLEGAGLRVS 360
Oy 523 TPVETWKGSKQKOS--VYTYIEENTTSFTWA-----FO-----RTPEHNASR 563
Db 361 VPLAAGNSYSESPSLYTFRLHEADYWPRLALSAFDOKLLHNLITAIKIRGYSEKSA 420
Oy 564 KITNDVAKIYINVTNMGV--ASYCRPCALEASDVSSCTSCAPAGYIDRDSGTCHS 620
Db 421 GHLDDV-----TITSVAPGAGVPAWVEESC-----SCPAGY--EEOFCER 458
Oy 621 CPPTNILKAHPYGVQACVPCGPGTKNNKIHSLCY-----LCVNDG--TFSRNTPTRTF 669
Db 459 CSPG--YRREP--GLGPRSPCVPCCTNG--HSETCPETGVC--DDRNTAGSHCKECS 511
Oy 670 NTFNSALNTV-----TLAGSPFTSKGLYFHHFTLSLCGNQGRKMSYCTDNTDLRI 723
Db 512 GYYGDTATGATLDCOPCCPGGSSCAVPRKTEVVCSTGCTGTRKCELCDD--AYFGD 569
Oy 724 PEGESGFSKITAYVCOAVIIPPEVTGYKAVSQPSIADRLIGVTTDMTLGI-----778
Db 570 PLGENAVAPRCLOCND--NIDPNAVGNCRQGTCLKIYNTAGFYCDKCKRGFGNPL 628
Oy 779 -TSPALFHEISLIPDVFYFRSDVYVOSCSGSHSTIRVCSPOKTVPGSLLPCTCS 837
Db 629 ATNPA-----DKCMAC-----HCNPYGVYVQOT--SCN 654
Oy 838 D--GTCDGCFHFLWESAACPL-----GSVADYHAIYSSCAVGIQKTYVWREP 885
Db 655 QVYTCQE--CLSHYTERDSCACERGFNLOSGRGRCRCDHALGST-----NGQCDINTGQ 708
Oy 886 KLCGSGISLPEQRYVTKITDF 907
Db 709 CECOPGV--GQRCDRCENHF 728

RESULT 11
O90021 PRELIMINARY; PRT; 719 AA.
AC 090021.
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Variant-specific surface protein.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
OC NCBI_TaxID=5741;
RN (1)
RP SEQUENCE FROM N.A.
RA Ey P.L., Darby J.M.;

```

"Comparative analysis of the VSP417 subfamily of variant-specific proteins in Giardia intestinalis.";
 Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF189719; AAF04387.1;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001744; Furin-like.
 DR InterPro: IPR005127; Giardia_VSP.
 DR Pfam; PF03302; VSP; 2.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00001; EGF-like; 1.
 DR SMART; SM00261; FU; 4.
 DR SEQUENCE 719 AA; 73888 MW; 83BE706BACE7977 CRC64;

Query Match

Best Local Similarity 3.7%; Score 202.5; DB 5; Length 719;
 Matches 133; Conservative 72; Mismatches 242; Indels 229; Gaps 41;

```

QY 39 TGTGPELHAKSESEHYETACDSGRMRAVPHITPGLCTSLDPYKGT---ECSEFS 94
DB 128 TAATKPSVISC-----DKTG---LITTAHT-----YKGEFCECSGS 163
QY 95 -----CNAGEFLDMKDOSC---KPCAEGRYSLGTGIRPDE---WDELPH 132
DB 164 ELTSGODGAKCTKCGASKYL--KDNACVDNAEAGKGYFGKPPAAGNKCITACTDSG 221
QY 133 GPASLSANNELDLSAESTGNCSTSSKWPVRDGT---APNDECTATLMYAVNLKOSGTV 189
DB 222 GTAGACECMYDSNTRKATCTT---DYLKKKADGTTECVAA-----NECDTQ 268
QY 190 NEEYYPDSITFEFFVONDOCPNADDSRMKKTTEKMEFHSVELRGNVLYWRTTA 249
DB 269 KCFYVNVST-----GNKVCSCADGAGLAVGTDAWK-----GV 303
QY 230 SVMTKVPKPVLRNIALTGAVATSECFPKRG---TYADKQSSF---CRLCPA 297
DB 304 DGCACKIRPADIN-----PTKDECKPGEISTDKTKCTSTAPPDPIENCAYC-- 353
QY 298 NSYSKGTSGHOCDDPKYSEKSSSCNVRA--CTDK-----DYFYTHACDANGETOL 350
DB 354 -----SEDRACEECNSNNY-----LPTMKCIDDCCKIGNIYTT-----SNANNKL 395
QY 351 MYWNAKPKICESDELEGAVKLPAAGVTHCPKCPNGPFKTN-----NSTCOPCPYGSYSN 404
DB 396 ICKCAVANKKE-----CENTGT--CKTCDGFGYKSSSECKADCSNCKTNGT--- 442
QY 405 GSCTRPACTEPAVGFEYKWMNTLPTNMETVLSGINFYKGMGWEMVAGDHITTAGA 464
DB 443 SADCCTKCLSG-----AVLKYGNNGTGTGCG---AGCATGTGAGA 478
QY 465 SDNDFMIL-----TLVYVGFPRPQSVNADPENKEVA--RITFEVETLCS--VNCLEYF 513
DB 479 CKTCGLIDGTISCECAVETEPPOGVCSTTVRAAATCKAGSVAKMCNSCTNGFLRM 538
QY 514 MGVNSRRTVPVETWKGSGKOSYTYIIEENTTSFTWAFORTPHEHSKRYTNDV--- 569
DB 539 NGGCVETTKFP-----GKS-----VCEEASAGDTQOKAPGVH-----LNNDLDTGCS 582
QY 570 --AKIYSIVV--TNMNG--VASYCRPALE-ASDVGSS--SCTSCPAGYVIDRDSGTCH 619
DB 583 PGCKTCTSTVYVACMEGYVKTSDSCACAGACATCTGCTACTDCTGTGY--KSGTTCV 640
QY 620 SCPPWTIIKAHQPYGVQACVPCGPGTKNNKTHSLCY 655
DB 641 SCTESNSDKTIT--GVANCAKACAP-PLNNKGSVLYCY 673

```

RESULT 12

O96J07 PRELIMINARY; PRT; 1316 AA.
 AC O96J07;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE CDNA FLJ1964 fis, clone PLACE400581, moderately similar to
 DE filipolepin 1 precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otaki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
 RA Nakamura H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
 RA Nakamura Y., Nagahari K., Masuno Y., Sasaki N.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

```

DR EMBL; AF027870; BAB5420.1;  

DR InterPro: IPR000152; Asx_hydroxyl.  

DR InterPro: IPR000561; EGF-like.  

DR InterPro: IPR000742; EGF-2.  

DR InterPro: IPR001881; EGF-Ca.  

DR InterPro: IPR003410; Hyalin.  

DR InterPro: IPR001759; Pentaxin.  

DR InterPro: IPR000436; Sush1_SCR_CCP.  

DR Pfam; PF00008; EGF; 6.  

DR Pfam; PF02494; HXR; 2.  

DR Pfam; PF00084; Sush1; 3.  

DR ProDom; PD002153; Pentaxin; 1.  

DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_5.  

DR PROSITE; PS00022; EGF_1; UNKNOWN_6.  

DR PROSITE; PS01186; EGF_2; UNKNOWN_8.  

DR PROSITE; PS01187; EGF_CA; UNKNOWN_5.  

DR SEQUENCE 1316 AA; 144524 MW; 80615BBA3AF00A5 CRC64;

```

Query Match 3.6%; Score 200.5; DB 4; Length 1316;
 Best Local Similarity 20.6%; Pred. No. 3.2e-07;
 Matches 197; Conservative 98; Mismatches 341; Indels 321; Gaps 52;

```

QY 47 HACKSEHYETACDSGRMRAVPHITPGLCTSLDPYKGTSECSFSCNAGEFLDMK-- 104
DB 562 HGFSFEMFYKARCDTDLMKRSEAFETTLGKWP-----SF-CSDAEDIDMWLE 612
QY 105 -DOSKPCAGRYSLGTGIRFDEMDELPHGF--ASLSANNELD--SAAESTGNCTS 156
DB 613 ENLTKKCYCLEYNDYENGFAIG-----PGWGAANRLDSTDDPLDYQETATSIGNAKS 667
QY 157 S---KWPVRDGT---IAFN-----TDECTATLMYAVNLKOSGTVNEFYPPDSITFE 203
DB 668 SRIKSNAPLSDIKIKLITFNITASVPLPDERNDTL-----E 702
QY 204 FVONDOCCPNADDSRMKKTTEKMEFHSVELRGNVLYWRTTAFSVWTKVPR----- 258
DB 703 WENOQRLLQTLFETITNKLKRLINKPMKSFOL--ASELIT--ADSNLSGTRKASPPCRPG 758
QY 259 -VLVRNIAT--GVAIYSE--CPCKPGTYADKQSSFFCKLCPANSYS---NKGETS 307
DB 759 SVLRGRMCVNCPLGYLYMLEHTECSCRISGYODEGQLCKLCPSCMYEYIHSRIISD 818
QY 308 CH-QCDPDKYSEKSSSCNVRAAPACDKDYFYTHACANGETOLMKWAPKICSEDLRG 366
DB 819 CKAOCKOGTSGSLFECESCPLOTYPKFGSRCLSC-----PENTSTVKKRG 866
QY 367 AVKLPAASVYKTHCPKCNNG--FEKTNSTCOPCPYGSYNSGD---CTRCP-AGTEPAVAF 421
DB 867 AVNISAGCV-----PCEBGFKFSRGLMPCHPCPRDXYQPNAGCAFCLACPFYGTTPFAG- 920
QY 422 EYKMMNTLPTNMETVLSGINFYKGMGWEMVAGDHIIYTAAGASDNDFMILTVYVGFPR 481
DB 921 -----SRITTECSRPS-----STFSAEBS-----VV-----P 943
QY 482 POSVADTENKEVARITF--VFETLCSV-----CELYFMVGVNSKRN----- 522
DB 944 PASIGHIKRHEITSQVHFECFENPCNHSNGTQOQGRGVYVCLCPUGY--TGAKCETDIDEC 1002

```



```

QY 523 TPV-----ETWKGSKGOSYTYIIEENTTTSTPTAFORTTPEASRKY 565
DB 1003 SPLCLNNGVCKDLVGEFICEBPGYTGOR-----CEEN-----1036
QY 566 TNDVAKIYSINTVNMNGVASY-CRPCALASDVSSCTSPAGYITDRSGTCHS--CP 622
DB 1037 INECSSSPCLNKGLICVDGVAGYRC-----TCVKGFEVLHCEETEVEECQSPCL 1084
QY 623 PNTLIKAHPYGVACVPCGPGCTKNK-----IHSLCYNDOCTFSRMTPTIFNYFSGALA 677
DB 1085 NNAVCEDDVGFLCKCPGFLGTGCKGVNDECLSQPCKNATCDG-----A 1131
QY 678 NTVTLAGPSPFTSKGLKFEHFTLSL-C-GNORRMSVCTDNVTLRIPEGESGFSKSI 734
DB 1132 NSFRLCAAGFTGS-----HCELINECQSNPCRNQATCVD-----EL 1169
QY 735 TAYVQAVIIPPEVTGYAGVSSQPVSLADRLIGVTTMDITDITSPALF-HLESIGIP 793
DB 1170 NSYSCKC---QPGFSGKRCE-----TQSTGFNLDFEYSGIYVLMGLPDLHAL 1218
QY 794 DVEFFYBSND-----VTOSCSGSRSTTIRVRCSPKTVPSLLPQTCSDGTCDGCF 846
DB 1219 TCTFWKSSDDMNITPISYANDNSDMT-----LLT-TDYN-----1255
QY 847 HFLW-----ESAACPLCSVADYH-AIVSSCVAGIQTYYWRERKLCSSGISTL 894
DB 1256 ---WVLVYNGRREKTCNPSVNDGRWHLIALTWTSANGIMK---VYIDGLSDGAGL 1306

RESULT 13
Q8T3A6 PRELIMINARY: PRT: 1045 AA.
AC 08T3A6;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Y47H9C.4C protein.
GN Y47H9C.4C.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_Taxid=6239;
OX
RN [1]
RN SEQUENCE FROM N.A.
RA Harris B.R.;
RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none.
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL: AL032657; CAD27615.1;
SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CRC64;

Query Match 3.6%; Score 197.5; DB 5; Length 1045;
Best Local Similarity 19.9%; Pred. No. 4.1e-07;
Matches 218; Conservative 83; Mismatches 329; Indels 467; Gaps 65;

```

```

DB 380 CSKTCIVENILMCPN---IGFCRCKRGFYGDNC---ELACSDSYGPNCEKO-----428
QY 224 TEKGEHSEHVELNRGNVLYWRTTAFSVTKYKPKVLYANLITAVAYISDECP-----277
DB 429 -----AMCDWN-----HASCEPNETGSCV 447
QY 278 CKRGYADKOGSFKLCIPANSYSNKGFTSCHQCDPDYKSEKSSCNVRPACTDKDYFV 337
DB 448 CKRG-----RTGKNKSEPCPLDFY---GPNCAHOC-----QCNR-----479
QY 338 THPAD-ANGETOAMYKMAKPKICSEDLGAVKLPASGVKTHCPNCPGFEFTNNSTCO- 395
DB 480 -GVCAGADKCCGCDGNGWCHR-CEH-----HCPA-----DTFGANCKE 516
QY 396 --PCPGYSYNSGD-----CTRCPAGTEPAVGFEYKMMNTLPTNMETVLSGINFEYKGM 448
DB 517 RCKCPKGI---GCDPIEGCT-CPAG-----LOGANDICGP 549
QY 449 TGMFVAGD--HITYAGASNDPMILLVYVGRPRPOSVADTENKVARITFEFTLCS 506
DB 550 EGSYGECKLHCKCVNCKCKETGECT-CQPGFFGSDCSTTCSGK-----YGESCE 600
QY 507 VNCLELYMVGVNSKNTNPTVETWK-----GSKG-----KOSTYIIEENTTTSTPTAF 553
DB 601 LSCPC-----SPASCCKTGKCLCPITKGVSCDCKCDPNFEGFLCQETVPS-----648
QY 554 ORTFHFAASKRYTNDVAKIYSINTVNMNGVASYCRPCALASDVSSCT-SCPAGYIYD 612
DB 649 -----PCASTDPKNGVCLSCP-----GSSGICHCHENCAGSYGD 683
QY 613 RDSGTC-----HSCPNTIILKAHPYGVACVPCGPGCTKNKIHSLCYNDCTFSRMTPT 667
DB 684 GCQOVCSADHGDPTT-----GECI-CEPG-----YHGKTCSEKCPDG 722
QY 668 TPNYNS-----ALANTVILAGPSPFTSKGLKFEHFTLSL-C-----GNORRMS 712
DB 723 KYGVGALDCKPKCASGSCDHINGLJCIPAGLEG-----ALCTRPCSAGFWNGCRQVC 776
QY 723 VCTNVDTLRIPEGSGFSKSTAYVQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTD 772
DB 777 RCTSEYKCKMNOTGECG-----CPA-----GFGQDRDCKCE-----808
QY 773 MTLDGITSPAEFLHLESIGIPDYIFETYSNDVYOSCS--SGR-----S 813
DB 809 ---DGYGPDICIKKCKCG-----TATSSCNVSGACHCHGFTGEFCHALCPE 854
QY 814 TTIYRCSPOKTVPGSLLPQTCSDGTCDGCFNHFLMESAAACPLCSVADYHAIYVSCVA 873
DB 855 STEFLKCSKECPKG-----CGDG-----YCDAAIGCC-----HYDOMSC-- 890
QY 874 GLOKTTYVWREPKLCSGISTLEPQRTVICTIDFWLKVAGISAGCTAILLVLCYFWKK 933
DB 891 GKAKQEFELNAGASTGLT-----WEPVLLIYALCGGLIAL--EYRN 933
QY 934 NOKLEYKSKLYMNAITLKDCLPADSCAIM--EGEVEDDL-----IFTSKN 979
DB 934 -----KYOK-----EKDPDPTVYSFHAKAPNNDGREGQNPPLYRSQVFPDSQAFSEN 981
QY 980 HSLGRSNHL--PPRGLL 994
DB 982 N-----GNHOGGPPNGLL 994

RESULT 14
Q8T3A7 PRELIMINARY: PRT: 1070 AA.
AC 08T3A7;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Y47H9C.4b protein.
GN Y47H9C.4B.

```

123 KYIGCALDPCKCASSSTCDHINGLCICPAGLEG-----ALCTRPCSAGFEWNGCROVC 776

interleukin-6; IL-6; PF00008; EGF; 9.

1. The first part of the document is a list of names and addresses of the members of the committee.

2. The second part of the document is a list of names and addresses of the members of the committee.

3. The third part of the document is a list of names and addresses of the members of the committee.

4. The fourth part of the document is a list of names and addresses of the members of the committee.

5. The fifth part of the document is a list of names and addresses of the members of the committee.

6. The sixth part of the document is a list of names and addresses of the members of the committee.

7. The seventh part of the document is a list of names and addresses of the members of the committee.

8. The eighth part of the document is a list of names and addresses of the members of the committee.

9. The ninth part of the document is a list of names and addresses of the members of the committee.

10. The tenth part of the document is a list of names and addresses of the members of the committee.

11. The eleventh part of the document is a list of names and addresses of the members of the committee.

12. The twelfth part of the document is a list of names and addresses of the members of the committee.

13. The thirteenth part of the document is a list of names and addresses of the members of the committee.

14. The fourteenth part of the document is a list of names and addresses of the members of the committee.

15. The fifteenth part of the document is a list of names and addresses of the members of the committee.

16. The sixteenth part of the document is a list of names and addresses of the members of the committee.

17. The seventeenth part of the document is a list of names and addresses of the members of the committee.

18. The eighteenth part of the document is a list of names and addresses of the members of the committee.

19. The nineteenth part of the document is a list of names and addresses of the members of the committee.

20. The twentieth part of the document is a list of names and addresses of the members of the committee.